

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 926 610 A1

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
30.06.1999 Bulletin 1999/26

(51) Int. Cl.⁶: G06F 17/60

(21) Application number: 97310368.2

(22) Date of filing: 19.12.1997

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE
Designated Extension States:
AL LT LV MK RO SI

• Hughes, Mischka
Stamford Bridge, York YO4 1JJ (GB)
• Jones, Michael William Dudleston
North Ferriby, East Yorkshire HU14 3NS (GB)

(71) Applicant: Infosis Group Ltd.
Market Weighton, York YO4 3AH (GB)

(74) Representative:
Driver, Virginia Rozanne et al
Page White & Farrer
54 Doughty Street
London WC1N 2LS (GB)

(72) Inventors:
• Geen, Nicholas
Hunmanby, North Yorkshire YO14 0PU (GB)

(54) Publication file conversion and display

(57) A computerized information display system extracts text data, lists of keywords, story rankings in order of story importance, and image maps identifying the location of stories from an input of publication files from a publisher. The system can generate a simultaneous display of a page image in which a story appears side-by-side with the text for the story when a particular story is selected, in order to allow a viewer can read the text while referring to the page image for visual cues about the text passage. The viewer can select a story from a displayed list of stories ranked in order of importance relative to other stories appearing on a page. The story rankings are derived based upon comparing one or more story importance indicators: location of the story on the page; size of type font of a headline associated with the story; size of type font associated with the story text; and size of text content for the story. The viewer can navigate to the text for a story on a displayed page by clicking in the story area on the page which is linked by image maps to the corresponding text passage. The viewer can also navigate to a text passage and page image by clicking on a keyword from a list of keywords extracted from the text input from the publisher. These computerized contextual display and image navigation tools allow the viewer a highly interactive experience with the publication. A method of converting a publication to an electronically viewable form is presented.

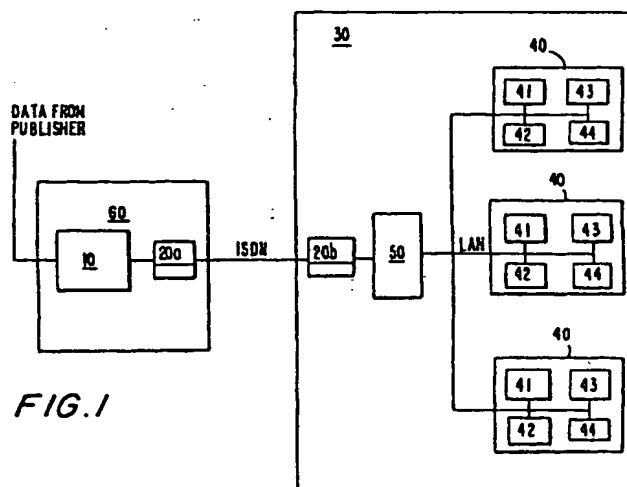


FIG. 1

EP 0 926 610 A1

Description

[0001] The present invention relates to a method and system for converting digital publications files into digital data, and the use of that data to generate a display on a computer system. Aspects of the invention relate to an information display system and more particularly to an information display system which provides for the simultaneous display of a graphical representation of a printed publication, or part of a publication, and text data appearing in the printed publication.

[0002] In today's society, particularly in the business community, it is a necessity to receive published information as quickly as possible. This is especially important for financial information. Thus, the desire to provide such information in electronic form has expanded rapidly in recent years.

[0003] In the United Kingdom, there are a number of suppliers of news information delivered electronically for on-screen or other media consumption. These can be segmented into a number of categories:

- (a) an electronic text feed of general and specific news items, and data where the only structure consists of headers detailing news category orders (e.g. Press Association);
- (b) an electronic text feed of news items addressing specific market sectors (e.g. Extel Finance);
- (c) an electronic text feed (not in real time) providing the textual information contained in previously published material. This information provided for archival and search activity as a primary facility (e.g. FT Profile).

[0004] The common component of these information provisions is their emphasis on editorial quantity, leaving the editorial and sub-editorial functions to the consumer. Essentially they are providers of a raw material to be used by the customer base as one of their ingredients for the production of their products, or as data for customers to filter to generate information for their own internal or external use. Thus, with this vast quantity of raw data provision with no relative importance attached to each of the individual news items, the user is forced to sift through irrelevant and/or unimportant information to discover their requirement. Additionally, the feeds are, in general, specifically objective rather than subjective.

[0005] A further disadvantage of this method of supplying information is that only text information can be provided. Although this text may be searchable or processable, as opposed to a graphical image or microfiche of the publication, it contains less information than the publication. In particular editorial information is lost. The foregoing problems of prior art information systems manifest the need for improvement. Specifically, there is a need for an information display system that can make use of or access information provided in publications

such as newspapers and magazines in real time thereby benefiting from the editorial experience of the publishers.

[0006] The present invention provides a screen based information display system which utilizes both the graphical images of pages of a printed publication as well as its text data. The present invention allows for the simultaneous display of an image of the pages of a publication and text data. It is not sufficient merely to provide a readable image of the pages of the publication as this only provides a microfiche representation. Whereas this allows the user to read the text, it does so at a representational level which does not give the overview perspective. The user "cannot see the wood for the trees", is a realistic analogy. The purpose of providing a simultaneous image of the publication is to allow the user to interpret the editorial importance that has been attached to articles, thereby allowing the user to benefit from the editorial experience of the publishers, as well as giving immediate access to the edited text.

[0007] The present invention allows for a user to select a passage of text comprising an article or story on the displayed page of the publication whereby the system of the present invention will simultaneously display the text of the passage adjacent to the image of the full page of the publication. This allows the user to clearly read the article if desired. In view of the small size of the image of the page of the publication the text is not clear and therefore it is highly advantageous to provide a clear copy of the text separately. The provision of the text separately also allows for further advantages of the present invention including allowing for identifier words such as company names to be clearly seen e.g. highlighted. The present invention provides for further information on the identifier word e.g. company information to be displayed, by the selection of the identifier word. The further information e.g. company reports, can then be displayed simultaneously with the image of the page of the publication.

[0008] A further feature of the present invention is that a list of contents of the pages of the publication can be displayed, wherein the list of contents for each page are displayed such that the passages of text (articles or stories) are listed in the order of importance which can be attached to them by the way in which they are formatted on the page of the publication by the editors. Thus, the list of contents for the publication provided by the present invention provides for an easy means for the important passages in the publication to be identified by a user. When a particular passage is identified which the user wishes to read, this can be selected and the text displayed along with the image of the page of the publication from which the text is taken.

[0009] The present invention is particularly applicable to business and financial publications such as newspapers. For example, in the United Kingdom, the London Evening Standard is published five times during a day with the financial information in each publication being

updated. Electronic data on each publication can be obtained rapidly from the publisher thereby allowing the information display system of the present invention to be updated rapidly in response to each new edition. The present invention thus removes the need for financial institutions to have to purchase multiple hard copies of the newspaper. Instead, the information can be provided electronically over a network to as many users in the institution as is required. Furthermore, the information provided is in a far more user friendly form than the original hard copy and reaches the user rapidly, even where the publication is printed some distance from the desired user, e.g. overseas.

[0010] For a better understanding of the present invention and to understand how the same may be brought into effect, reference will now be made, by way of example, to the Figures, in which:

Figure 1 illustrates an exemplary embodiment of a system for implementing the present invention;

Figure 2 illustrates a display generated during the operation of the system illustrated in Figure 1;

Figure 3 illustrates another display generated by the system of Figure 1;

Figure 4 illustrates a further display generated by the system of Figure 1;

Figure 5 illustrates yet another display generated by the system of Figure 1;

Figure 6 illustrates a still further display generated by the system of Figure 1;

Figure 7 illustrates another display generated by the system of Figure 1.

Figure 8 depicts an overall process flow for converting raw publisher input to simultaneous text/image display of a publication.

Figure 9a is a flow chart illustrating one method of converting the publication files supplied by the publisher into a data structure.

Figure 9b is a flow chart of an example of an importance-determining model for ordering a list of stories on a page by relative importance.

Figure 10 illustrates the use of keyword lists to navigate to the text passage and page image containing the keywords.

Figure 11 illustrates the use of image maps to navigate to the text passage on a page image containing the imaged article.

[0011] Referring now to the drawings, and initially to Figure 1, there is illustrated an exemplary embodiment of a system for implementing the present invention. Data is received from the publisher in electronic form by the central storage and processor unit 10. Whilst it is highly desirable that the data be obtained from the publisher in electronic form, it is not essential to the principle of the present invention. Any means of providing images of the publication and separate text data will suffice.

[0012] Within the central storage and processor unit 10, portions of each page of the image which correlate to passages of text are defined and the defined portions are correlated with the passages of text. A list of contents for the passages of text is then generated by selecting the headings from each passage of text and ordering these in order of importance which can be attached to each passage of text by studying the image of the page of the publication. For instance, where an article has the largest heading in the publication of a newspaper, clearly this is the most important story of that page. Similarly, if an article has the smallest heading, this is the least important story on that page and is thus placed at the end of the list of contents for that page. Once the list of contents is generated, this is stored for later assimilation into the invention. A detailed description of an exemplary process for the above is provided in the section below entitled "Ordering Text Passages and Generating A List of Contents".

[0013] The image received from the publishers or obtained from the publication requires enhancement of visual quality and therefore in an embodiment of the present invention the received image is sharpened to improve the definition and therefore make it clearer when displayed. A detailed description of an exemplary process for the above is provided in the section below entitled "Enhancing Visual Quality Of Page Images".

[0014] Within the text information there will be certain words such as company names which can serve as identifier words for which the central storage and processing unit 10 has further information which can be made available to the user. Therefore, the text data which is received from the publisher is searched and compared with known identifier words such as company names. The identified identifier words are then flagged in the text and are also entered into an index which is then stored for later assimilation into the invention. A detailed description of an exemplary process for the above is provided in the section below entitled "Generating a List of Identified Words".

[0015] Additionally, within the page images and text information, there will be stock market equity price information from a variety of Stock Exchanges around the world, together with the price movement on those equities. These prices and price movements will be those standing at the time of the publication of the newspaper. Within the central storage and processing unit 10 there is additional information on many equity companies, including the current real time price of these equities.

[0016] Therefore, the text data and the images which are received from the publisher is searched and the particular companies used within the publication identified and the further information and real time price data within the central storage and processing unit 10 can be made available for the user when assimilated into the invention. A detailed description of an exemplary process for the above is provided in the section below entitled "Linking Identified Information With Further Data".

[0017] Thus, the information that is available from the central storage and processing unit 10 is a series of images of pages of the publication, the text corresponding to the articles or passages in the publication, which passages of text have been correlated with the particular portions in the image, a list of contents of the passages of text listed in order of importance for each page, an identifier word index identifying the words e.g. company names in the text for which further information is available, and further information on the identifier words e.g. company prospectuses or statistical information together with real time equity price and other information on the companies within the market price pages of the newspaper.

[0018] The series of steps to compile this data will be carried out for each publication. Thus, in the case of a newspaper for which there are several publications in a day, this process must be carried out for each publication as quickly as possible in order that the information can be made available to users without delay.

[0019] The central storage and processing unit 10 can then communicate the stored information using a communication link 20a and 20b to a single user or group of users 30 such as a financial institution. In Figure 1 the communication link is a high speed ISDN telephone line. However, any form of communication can be used such as internet, cable, satellite or radio. Typically, in such an institution where a plurality of users require information, each user will be provided with a personal computer or terminal 40 which is connected via a local area network (LAN) to a central processor which is for instance a file server 50 which receives the information via the communication link 20a and 20b from the central storage and processing unit 10. A detailed description of a specific implementation for the above is provided in the section below entitled "Output Of Information System Displays To Users".

[0020] Thus, each personal computer or terminal 40 has access to all the information available from the central storage and processing unit 10 at the remote location 60. The central storage and processing unit temporarily stores the information as a digital data structure before transmission in a memory and each personal computer or terminal 40 stores the information as a digital data structure on reception, in a memory. Each personal computer or terminal 40 comprises a central processor unit 41, a memory 42, a display 43 and an input device 44 such as a keyboard and/or a pointing device such as mouse or tracker ball.

[0021] In order to make the interface of the computer system with the user as easy as possible, according to one embodiment of the present invention, the software utilized in the personal computers 40 operates on the basis of displayed icons which illustrate and control the running of options. The icons are selectable and operable by a pointing device such as a mouse. Each of the icons forms part of a control item, the other part being a link to a text file, page image file or other displayable

information. Selecting the icon activates the control item which uses the link to retrieve and display the displayable information. However, the present invention is not limited to the use of a cursor movement device such as a mouse and instead any means of inputting selections and commands, e.g. a keyboard falls within the scope of the present invention.

[0022] Referring now to Figure 2, there is disclosed an image which is displayed when the embodiment of Figure 1 is in operation. On one half of the screen there is a page preview of a page of the publication and the page number (page 33) is indicated as well as the title in the top left-hand part of the display. In the left-hand part of the screen there is displayed a list of contents for the pages of the publication listed by page number and for each page number the articles are listed in order of importance. The list of contents can be scrolled up or down and the next and previous pages of the publication shown on the page preview on the right-hand side of the display can be selected, although in this Figure there is no previous page since this publication has no pages prior to page 33. At the top of the screen icons are provided to allow either the next or previous page image to be viewed, by their selection. Each icon is part of a control item, the other part being a link to a page image file of the next or previous publication page respectively. Selecting an icon activates the control item which uses the link to retrieve and display the relevant page image file. The display of the list of contents is selectable by selecting the contents option at the top left-hand part of the display by moving the cursor and depressing the mouse button i.e. "clicking" on that icon. This activates a control item and the link to the list of contents file is used to retrieve and display the list of contents. It is also possible to select an article on a page to be displayed by moving the cursor to point out the article listed in the contents and clicking on it. This will activate a control item and the text file will be retrieved via a link to display the text of the article in the left-hand part of the display whilst in the right-hand part of the display the image of the page on which the article appears will be shown.

[0023] Referring now to Figure 3, in this display the article headed "German cheer for shares and bonds" has been selected by moving the cursor to the portion of the image and clicking on it. The image is then highlighted by coloured border or indicated by a web browser with an "active" icon, e.g., a pointing finger as used in the Netscape Navigator™ browser, while on the left-hand side of the screen the text of the article is displayed. The text displayed can be scrolled up or down in a conventional manner. At the top left-hand part of the screen icons are provided to allow either the previous or next story to be selected. Each icon is part of a control item, the other part being a link to the text file of previous or next story respectively. Selecting the icon activates the control item which uses the link to retrieve and display the relevant text file. In the display of Figure 3

there is no previous story since the selected story or article is the first of the publication.

[0024] Within the story or article there may appear references to companies. When such references occur, these are highlighted in the text and a user can select to view further information on that company by moving the cursor to the highlighted text acting as an identifier word and clicking on it. The highlighted text (identifier word) acts as an icon and forms part of a control item, the other part of the control item being a link to further information. Clicking on the identifier word activates the control item and causes the retrieval and display of further information in at least the left-hand part of the screen. Such further information can for example be a company prospectus or company report.

[0025] Figures 2 and 3 also show in a bottom left-hand part of the display that the icon "find" is available. Next to this, it is possible to enter a string of text which the user wishes to find within the text of the publication, once the text string is entered in the string entry field and the "find" icon is activated. Once the text string is found within the text, the article in which it appears is displayed in the left-hand part of the display together with the page on which it appears in the right-hand part of the display. The text string within the article is highlighted.

[0026] The display in this embodiment of the present invention is provided with the ability to select a company index. This is provided for in the bottom left-hand corner of the screen as a "Company" icon. This icon forms part of a control item, the other part being a link to a company index. Selecting the icon activates the control item which uses the link to retrieve and display the company index. When the icon is selected, the display of Figure 4 is generated. In Figure 4 in the left-hand half of the screen, an index of the companies referred to in the publication is given. By moving the cursor to a particular company name and clicking on it, text is displayed on the left-hand side in which the first mention of the company name occurs and on the right-hand side of the display there is displayed the associated page of the publication. Where there are a number of publications per day, the index of companies can indicate next to a company name the publication number during the day in which there is a mention of that company. This gives further information on the number of times a company is mentioned in the publications throughout the day and thus gives an indication on the importance of the activities involving that company.

[0027] Figure 5 illustrates a display of financial information in the publication. In the page of the publication the financial sector can be selected and under that sector the financial information on the companies can be displayed. The financial information available can be far and above what is available in the publication since further financial information is available and can be obtained from other sources and collated in the central storage and processing unit to make it available to

users.

[0028] Referring now to Figure 6, this illustrates a further display wherein the text on the left-hand side of the display not only includes highlighted company names but also includes a processed image which originated from the image portion on the page preview on the right-hand side of the display under the heading "Footsie Reels from Iraqi Shockwaves". The processed image of the graph can be manipulated by the user. Further information over and above what is available from the publication can be included in such processed images. Such further information can be made available from alternative sources and can be combined within the central storage and processor unit 10.

[0029] Referring now to Figure 7, this illustrates a further display wherein further information over and above what is available in the publication is selected and displayed. In the page preview on the right-hand side of the display there is an advert for a computer manufacturer. When the cursor is moved to this portion of the image and clicked on, further information which comprises further advertisement information is displayed in the left-hand side of the display.

[0030] When the option of requesting further information is selected, the software moves out of the current application and into another application containing the required additional information. Such further information can take any form such as graphical, textual and video information, thus allowing the present invention to operate as a multi-media software system. Thus, the information display system of the present invention, by providing both a graphical image of a publication and the text data, acts as a gateway through the publication into a vast array of further information which can be made available to the user via the central storage and processing unit 10.

[0031] A specific implementation of the information display system of the present invention is described in detail below for a given example of an electronic publication. For the described implementation, an overall process for converting raw publisher input to simultaneous text/image display of a publication is depicted in Figure 8. The conversion process includes the steps of extracting text data and related graphic images and processing page images (indicated at blocks 81, 82, 83 and in Figure 9a) from the publisher's raw input (80), generating a list of contents (84) and a list of company names (85), authoring the simultaneous text/image displays of the publication (86, 87, 88, 89), and providing the information display system as an output to server/users on a network (90). It is to be understood that the invention is not limited to the described implementation, and may be implemented in any equivalent manner using the disclosed principles of the invention.

Ordering Text Passages and Generating A List of Contents

[0032] The information display system requires two basic types of input, text data and images of pages of a publication. The publisher typically provides the new data for an issue of a publication in digital electronic form, for example, as publication files, such as Quark XPress™ files or as PDF (Portable Document Format) files used in page make-up systems and readers offered by Adobe Systems, Inc., of Boston, Massachusetts. Text is extracted from the Quark Xpress™ or PDF files using the built-in functionality of the page make-up program and classified as data entries for storage and retrieval from a data base as digital text data files. The text of each story has a corresponding digital text data file. The page images can be created from Quark Xpress™ files by first producing EPS (encapsulated Postscript) files, or from the PDF files by first converting them into EPS files. Each page image is stored as a digital page image file. The processing of the publication files to create page image files can be automated as described in the section "Enhancing Visual Quality of Page Images".

[0033] The publication files are in a format suitable for editing the publication document or printing the publication document. The publication document consists of a number of pages each of which contains one or more stories. Each story has at least a headline and a text portion and may in addition have an associated picture. A representation of each page in the published document is produced from the publication files and stored as digital page image files as described in the section entitled "Enhancing Visual Quality of Page Images". Each page image file is associated with the page of the publication on which it appears and can be used to reproduce the image of the page on a visual display unit. Each page image file may be a bitmap of one page of the publication. The publication files are also processed to extract from each page, the stories which are on that page and for each of those stories the headline, text portion and any pictures associated with that story. This process of extraction may be achieved in any one of three ways.

[0034] According to a first method, the publication files contain additional format data which identifies where each story is positioned within each page and where each story starts and ends, where each story's headline starts and ends and the font size of the text used within the headline; where the body of text making up the story starts and ends and the font size of text; and where any picture associated with the story is placed within the page. Such format data is not observable in the image of the published document but describes or controls the format of the published document. A digital processor operates on the digital publication files to extract this additional format data and to create data files for each story including: a headline text file containing informa-

tion identifying at least the text content of the headline and the headline font size; a story text file containing the text of the story and information identifying the text font size; a picture file containing information sufficient to reproduce a picture associated with the story such as a bit map image of the picture; a picture position file indicating where the associated picture is placed in relation to the text; and a story position file indicating the limit of the boundary of the story in the page image.

[0035] A second method can be used in the absence of additional format data in the publication files. In this method, format data can be derived from the publication files by a digital processor. The first stage is the determination of the number of separate stories in a page of the publication. This is achieved by using the format used to divide individual stories, which may be lines or blank margins for example, to identify the boundary of each story. The processor having identified the number of stories on the page takes each story in turn, and for each story produces data files including: a story position file, a headline text file, a story text file, and if appropriate a picture file and picture position file. The story position file is produced by identifying the limit of the boundary of the story within the page image. The headline text file is produced by identifying the text within the boundary of the story which has the largest font size. The headline text file stores the text of the headline and information identifying the size of the font. The processor may then assign any remaining text within the body of the story to the story text file and also store information identifying the font size of that text. The processor may then identify pictures within the boundary of the story and create a picture file containing a bit map image of the picture and a picture position file storing information identifying where the picture was positioned within the story. The processor then goes through the same process for each of the stories on the page and for each page within the publication.

[0036] According to a third method, an operator creates the data files including the headline text file, the story text file, the story position file and any picture files and picture position files by selecting areas of the image of the publication displayed on a visual display unit using a cursor control device. The limit of the boundary of the story in the page image is first selected and this information is stored in the story position file. The operator then selects the headline of the story and a headline text file is created which stores text data and information identifying the font size of the text. The operator then selects any pictures in the story and the bit mapped image of the picture is stored in a picture file with the positioning of the picture within the story being stored within a picture position file. The digital processor then stores the remaining text within the boundary of the story which has not previously been selected and information identifying the font size of the text in a story text file.

[0037] A data structure is now produced which inter-

links the various components of the publication including the data files and the page image files. A RECORD is created for each story on each page. Each RECORD has a one-to-one correspondence with a story on a page. Each RECORD contains a number of fields which associate the RECORD to the data files and page image files of its corresponding story. The first field is a POINTER to the headline text file of the corresponding story, the second field is a POINTER to the story text file of the corresponding story, the third field is a POINTER to any picture file associated with the corresponding story, the fourth field is a POINTER to the picture position file associated with the corresponding story and the fifth field is a POINTER to the story position file of the corresponding story. Consequently, the digital processor parses the publication into pages thence into stories and each story into its component items such as headlines, text portions and pictures. It produces a data structure consisting of a plurality of data files, page image files and RECORDS which interlinks the components of the publication, and from which the publication can be recreated in different electronic formats.

[0038] The RECORDS are now indexed. Each RECORD is indexed by a page number (page_no) and a story number (story_no). For a particular RECORD, the page number indicates the page of the publication on which the corresponding story appears, and the story number identifies the corresponding story amongst the other stories on that page. Consequently, the combination of story number and page number uniquely identifies each RECORD and its corresponding story. The story number (story_no) is used not only to uniquely identify a story on a page but is also used to indicate the relevant importance of a story in comparison to the other stories on a page. The most important story on a page will be assigned a story number 1 with the value of the story number increasing as the importance of the story diminishes.

[0039] The story number can be assigned on the basis of operator judgement or by digital processing. Each RECORD contains fields having POINTERS to data files containing all the information associated with a story. Each of the RECORDS corresponding to the stories on a particular page can be processed to determine the relative importance of the stories on that page. For each of the RECORDS, the processor accesses the associated headline text file, the story text file and story position file. From these files, the processor can, in relation to each of the stories, determine the positioning of the stories on the page relative to one another, determine the headline font sizes relative to each other, and determine the story text font sizes relative to each other. On the basis of this information, the processor can order the stories in relative importance. Generally, any story that continues from a previous page will be given the highest relative importance and the remaining stories will be rated in dependence upon the font size of their headlines with any two stories having the same font size

for the headlines being differentiated on the basis of the position of the story within the page and the font size of the text in the body of the story. It will be appreciated that the model used to weight the relative importance of the various types of format information will depend upon the particular editorial style of the publication and a different model with different weights applied to the different types of format information can be used for different publications. A flow chart of an example of a model suitable for determining the relative importance of a story within a page and creating a list of the stories on a page ordered in terms of their relative importance is shown in Figure 9b.

[0040] The process of creating the data structure includes extracting the data files and page image files and creating records is illustrated in Figure 9a and steps 81,82,83 of Figure 8.

[0041] Once all the stories have been indexed through RECORDS the data structure is processed by means of a digital processor to produce output files, or an output signal which can be used by an end user to access the information stored within the data structure and hence within the publication and display that information on a visual display unit (VDU). The end user will be able to view via the page image files accurate representations of the pages of the publication. The end user will also be able to view the text of each story in a clear form via the story text files. In addition the VDU which the end user is using will have a series of icons on the screen which can be selected by using a pointing device such as a mouse. If an icon is selected the end user is able to navigate through the publication.

[0042] According to one example the digital processor processes the data structure and produces an output in a HTML format suitable for use in an end user's browser software such as Netscape Navigator™ or Internet Explorer™. The processing of the data structure transforms the data structure into a code which, on an end user's machine produces an electronic publication having actuatable control items. The control items comprise a visual symbol on the VDU of the end user's machine, such as a word icon, and a link from the visual symbol to other information. In HTML this may be achieved by creating an anchor and a hyperlink. Actuating the visual symbol using a pointing device accesses the other information and enables its display on the VDU. Consequently, when the code is loaded into a computer by an end user a display as illustrated in Figures 2 to 7 is produced having a page preview produced from the page image files, a clear text portion produced from the data files and a number of icons for navigating through the publication produced by processing the data structure. These icons include previous/next story icons, previous/next page icons and a contents icon.

[0043] The next/previous page icon allows the end user to move through the pages of the publication. If the next page icon is selected the page image file associated with the page following that being currently viewed

is loaded for viewing by the user. If the previous page icon is selected the page image file associated with the previous page of the publication is loaded for viewing by the end user.

[0044] The next/previous story icons allow the end user to navigate through the stories on a particular page. Selecting the next story icon displays, in clear text format, the story with the next, lower level of importance on the page. This is equivalent to accessing the story corresponding to a RECORD having the same page number but with a story number one greater than the RECORD corresponding to the story currently displayed on the screen. Selecting the previous story icon displays, in clear text format, the story with the next, higher level of importance, on the page. This is equivalent to accessing the story corresponding to the RECORD having the same page number but having a story number one less than the story number of the RECORD corresponding to the story currently being viewed.

[0045] Selection of the table of contents icon displays an ordered list of titles equivalent to ordering the RECORDS firstly according to their associated page number, ordering those RECORDS with the same page number according to their story number, and then accessing through the first field of each RECORD the headline text file for each story and displaying a list of headlines in the same order as the RECORDS. Consequently, a table of contents, as illustrated in Figure 2, can be produced which illustrates the titles for each page of the publication, ordered in dependence on their relative importance. Each title on the page of the table of contents is an anchor for interactive linking to the story in clear text format and/or page view format.

[0046] When a particular page image file is loaded, it is possible for the end user to locate the cursor over a particular story on the page image and select that story. The story text file associated with the selected story will be loaded and the story displayed in a clear text format as illustrated in Figure 3. When the selection is made the page number associated with the page image file currently being viewed is known, and the location of the cursor within the page image when the selection was made is known. The display of the selected story is equivalent to searching the RECORDS to select the one which is associated with the correct page number and which has a POINTER in its fifth field pointing to the story position file defining the area in which the cursor was positioned when the selection was made, and displaying the text data and other information of the selected RECORD on the VDU.

[0047] Once the text data have been extracted and the text passages have been assigned page and story numbers, a list of the contents of the publication. In the information display system output, the contents list can be called up for display on the left side of the screen simultaneously with a page image on the right side of the screen (see Figure 2), to serve as a guide to users of

the contents on the current page and on the preceding and succeeding pages of the publication. Each entry in the list is a selectable "icon" forming a link to the page on which the entry appears.

[0048] An example of a program for generating a contents list is given in Appendix 1. The contents list is generated from the text headlines or subheadings associated with the stories ordered by page numbers and story numbers. The generated headline index file is output for use in generating the simultaneous text/image displays for the given publication ("London Evening Standard Business Day" in this example).

[0049] As shown in Fig. 11, the text passages (text stories) are linked by image maps IMx to the locations A-i of the corresponding stories on the page images. Each of the locations A-i of the story areas on the page images is similar to an icon in that it forms part of a control item, the other part being a link to the corresponding text passage (story). Clicking on a story area activates a control item which uses the link to retrieve and display the text file corresponding to the story in the selected story area. In the information display system output, the text passage can be called up for display on the left side of the screen simultaneously with display of the page image with story highlighted on the right side of the screen (see Figure 3), to allow users to view the text in detail and interact with any linkages therein together with the contextual and editorial cues provided by the page image.

[0050] Image maps are used so that the story area acts like an icon or selectable button. The image maps can be created using, for example, mapping software such as one designated Web Map, which is available as shareware and stored as digital image map files. Typically, a rectangle or other shape is overlaid on the processed image by an operator who links the pixels within the shape on the image map template with a page number and story number in the database. This can be done by indexing the text file to the pixel group using a corresponding file naming convention, e.g., a "P1S2" suffix for the text file corresponding to the article area delineated on page #1 as story #2. The text files are read into the database which then stores the coordinates of the pixels contained in the map file with the record for that story. It does this by using the file name to identify the corresponding record in the database, in this case, the text record for page #1, story #2. A field in the database is updated to contain the indexing information.

[0051] The input data conversion process can include the extraction of other pictures and graphics appearing in the page images which are related to the text passages, or of cartoons, advertisements, and other graphics which may be desirable to display in their own right simultaneously with the page images. The graphics images are extracted from the EPS or PDF files into individual graphics files using standard graphics editing tools, e.g., the Adobe Illustrator™ system. Graphics

related to the stories, such as a photo of the subject of a story or a headshot of a contributing columnist, are indexed in the database to the stories by page numbers and story numbers. Besides extracting the Postscript images in the manner described above, sufficient quality can also be obtained by using "screen dumps" of the page make-up files themselves and separating the bit-mapped components. This can be achieved, for example, using the Adobe Photoshop™ system. Standalone graphics can be linked to their locations in the page images using a control item and the image mapping described above.

[0052] In the information system display output, story-related graphic images appearing in the page images can be retrieved, manipulated, and displayed on the left side of the screen in a window adjacent to the text passage (see Figure 6). Standalone graphic images can be called up for display on the left side of the screen by a mouse click, or can be used to trigger an external retrieval process resulting in display of a linked graphics file, such as an advertisement (the "Dell" logo linked to the advertisement in Figure 7), an externally retrieved output (the up-to-date stock performance graph in Figure 5), or the display of an associated text passage.

[0053] The extracted text data, list of contents, image maps, and extracted graphic images are stored in the database along with the processed page images. The database thus contains an ordered, structured, and mapped version of all text and related graphics components linked to their positions in the page images.

Generating a List of Identified Words

[0054] A list of important identifier words appearing in the pages of the publication can also be generated from the extracted text data. Important identifier words can include the names of companies, important persons, well-known products, media programs, etc., which are reported on in the publication. In the information display system output, a list of company names reported on in the publication can be called up for display on the left side of the screen (see Figure 4). A click or entry of a selected company name will result in a display of the page image and highlighted story in which the company name appears on the right side of the screen, and the corresponding text passage on the left side of the screen (see Figure 6). Similarly, display of a text passage with important company names highlighted therein allows a user to click on the highlighted name or word and call up another display of further information on the company.

[0055] Keywords are often designated in text by the publisher, for example, using specialized type fonts such as using bold font for company names or using italics for author's names or publication references. This designation in the text constitutes format data and provides a convenient way to identify keywords from the publisher's input. In the example illustrated in Appendix

1, company names in the input from the publisher are highlighted by tags for bold type font. Thus, a list of company names can be generated using a digital processor to parse the digital text data and extract the names delimited by the bold tags into a company index file. The company names on the list are then indexed to the page numbers and story numbers where they appear in the page images, as well as by their text positions as delimited by the bold tags in the text passages. Each keyword text position is consequently indexed to a page number and a story number, and a link is formed between the text position and the story (text and/or image) in which the keyword appears. The text position and link form a control item which is activated by clicking on the text position. Activation of the control item causes the story to be retrieved and displayed. Indexing by their text positions allows the company names to be highlighted in the text displays and defined as control items having anchors for interactive linkages to further information. The resulting company index file is stored in the database for the simultaneous text/image information display system. The company names may also be added to a company name library list which is cumulated over time. In this manner, extensive keyword lists can be developed, and may be used for alternative methods to automated parsing of keywords.

[0056] One alternative method for generating the list of important identifier words (keywords) is to use a digital processor to search for text strings in the extracted digital text data which match entries in stored library lists of known company names, important person names, product names, media names, etc. The library lists can be updated from the electronic files processed and/or by manual input of an operator when a new keyword is recognized. When important identifier words are identified in the text passages, the digital processor adds the names to the keyword list, indexes the names to their page numbers and story numbers indicates the names to the positions of the words in the text passages, and creates a link between the name and the story in which the name appears. Keywords can also be added manually to the keyword list by the operator. The keyword lists serve as a powerful method of navigation to the covering stories in the simultaneous text/image information display system. Figure 10 illustrates the selection of a keyword in a keyword lists to navigate to the text passage and page image containing the keyword. The image maps also provide the ability to navigate among stories on a page and call up the corresponding text passages by clicking on the mapped areas of the stories.

Enhancing Visual Quality Of Page Images

[0057] Along with the above, the page images are processed by processing the encapsulated postscript files of the publisher's input files to form 72 dpi bit-mapped page images or any other resolution appropri-

ate to the intended output medium. The Internet, for example, usually requires images in GIF (Graphic Interchange Format) where the file sizes are minimized to enhance the speed of download. Optimized palettes are used to also minimize file size and increase visual quality. The image files are manipulated using bit-map processing software, such as Adobe Photoshop™ or Debabaliser™ software, to produce page images that are visually enhanced and/or data compressed to be acceptable in quality and reasonably small in file size. Scripts can be written to batch process the EPS or PDF files into appropriate page image files in a completely automated process. These scripts call a set of routines commonly used in image manipulation software such as Photoshop™ or Debabaliser™ software.

Linking Identified Information With Further Data

[0058] Further information is stored in the database by establishing indexes to its parent page/story number. "Regular" features, e.g., where the same page/story is always written by the same author, and may include a picture of the author, can be added automatically as a default by the database. Others are identified by an operator who may use a pull down menu of regular features or may insert the name manually. The naming convention of "P1S2G3" may be used typically indicating the graphics #3 connected to page #1, story #2.

[0059] A typical picture or graphic image would appear somewhere before the main body of text. Its position is indicated in the database by a number which instructs the database to output the link to this item after the appropriate numbered text item. Where a picture/graphic element is desired to be presented within the main body of the text, a convention of "[n]" (number within square brackets) may be used for the graphics number to instruct the output stage of the database to substitute this sequence for the appropriate longer form of the graphics name. This is designed to avoid operator error in miskeying longer sequences of characters than necessary in these manual operations.

[0060] Linkages to external data sources, i.e., external to the original publication, is typically achieved by linking to a predetermined set of hooks in the database. For example, a share price for a company identified by the keyword indexing process can be obtained using the official company name or stock exchange symbol stored in the database. After looking up the unique identifier name in the database, the system performs a share price look-up procedure with an external data source, and returns the retrieved share price for use in the display system.

Output Of Information System Display To User

[0061] When the conversion of the publisher's input data in the database has been completed, a software routine in the display system creates a sequence of files

containing the desired sequence and style of displays, linkages to both internal and external data, and other interactive functions for the information display system, as illustrated in Appendix 2. The linkages between story areas and related graphic images, text passages, keywords in the text passages, and image maps for the page images defined in the data conversion stage are used to define display buttons, highlighted stories, highlighted words, and linked displays in the display authoring stage. In accordance with the invention, examples of displays of text passages simultaneously with the page images providing contextual cues for the text passages to the viewer are shown in Figures 2 - 6. The resulting processed files constitute a digital data structure viewable using Web browsing software such as Netscape™ Navigator or Microsoft™ Internet Explorer on a file server running server software such as Novella™ Netware, Appleshare™, or Windows NT™ Server software. The digital data structure can also be uploaded onto a Web server running Netscape™ Server, Microsoft™ Information Server, or Apache™ server software. Once in the database, given the structuring of the data as described, the created files can be converted to a digital data structure in one of many possible formats and stored in a memory.

[0062] The Web-viewable files (digital data structure) are transmitted from the memory of the processing unit 10 to an intended server using suitable transmission software which first identifies files as either new or unchanged from a previous transmission. The Web-viewable files (digital data structure) are stored in a memory in the intended server. If the files are changed, they are compressed into a single file and transmitted over ISDN, PSTN, or leased line to the receiving server. The receiving server unpacks the compressed file into its components and copies them into the appropriate place on the user's server. This approach is used for efficiency where multiple destination types may be required. It does not matter whether the user's server is a true file server or whether it is a Web server. The transmission software can also be configured to compress all data all of the time. Alternate software routines, such as the "TAR" function used in the UNIX™ operating system, can be used to combine all files for transmission to a remote server. A simple UNIX™ script can be used to scan for these files and decompress them with the "UNTAR" function and copy them to the appropriate directory.

[0063] In an information display system configured for an intranet as shown in Figure 1, each user is provided with a personal computer linked to a central file server to provide the necessary information. The personal computer can be a Macintosh System 7 with a 256 colour video screen or an IBM (Registered Trade Mark) compatible 486 based computer having at least 8 M Bytes of RAM and a 256 colour video screen. The information display system may also be configured as a server for the Internet to which a universe of users and

server nodes may have access.

[0064] The server database for the information display system of the invention has the display content assembled into formats suitable for the medium to which it is targeted. For example, HTML is usually used where the output medium is the World Wide Web. However, the database can also process the system output files in other structured formats, for example, the Bloomberg™ real-time display system. The common principle is that the linkages between text passages, keywords, graphics, and page images are ordered based upon their assigned priorities and/or locations on the page images.

[0065] In addition to the solution of providing simultaneous displays for communicating familiar visual contextual and editorial information to the reader of text passages in the publication, the provision of the system output files in a readily viewable HTML format provides certain other advantages over simply viewing the publisher's publication file in PDF format with a plug-in or other formats not supported on the Internet currently. A PDF file is published as a single large file. Although there are advanced download technologies available, a PDF file typically takes longer to download than an HTML file, which contains many smaller files. An HTML file is also simpler to edit, integrates seamlessly with other Web technologies, and can provide access to further information from other active processes and databases dynamically.

[0066] The information display system of the present invention may be modified and extended in other ways. For example, since the text data are extracted from the publisher's input and maintained in the system database, the text data can be readily searched by any search engine to find target stories, names, and references and to retrieve the publication pages containing them. The processed information in the form of the prioritization of stories by importance and keyword lists can be used to assist with conducting high quality searches with high efficiency. Thus, the published issues can be converted to a data resource that is fully accessible and searchable by external users.

[0067] The processing of publisher's input into system output files and authoring of linkages between text passages, keywords, graphics, and page images can be developed further for fully automated processing. Batch processing scripts can be developed for automatically extracting text data, graphics images, and keywords, generating image maps, and updating system library files. Stories may be tagged in the database in such a way that advertisements handled by the system will be changed as different stories are selected. This would allow customizing of advertising opportunities by associating different story types with different advertisements.

[0068] The processed information obtained in the present invention may also be used in other ways to provide further advantages. For example, the image maps defining the story areas for the page images may be

used with the original PDF files to provide the capability for enhanced functions. The image map could be overlaid on the PDF file itself and allow a simple click to simultaneously view the chosen text similar to the display result described previously. Additionally, the PDF file would retain its zoom capabilities inherent in the Acrobat™ reader software from Adobe Systems, Inc. Clicking on the story area of an image map can be used to trigger an internal process, such as zooming in or out on a page view, or an external process, such as connecting to a related database of supporting information.

[0069] If the headlines are too small to read on a given display of a bitmapped page, a mouse roll-over can pull up headlines or captions in a pop-up box. Another possibility is to create a floating window which contains the page image. This could develop into an information tool, taking advantage of push technologies and broadcast methodologies using the JAVA language as its development platform rather than HTML, wherein a click can pull up a simultaneous display of background material in another display window.

[0070] It should be understood that the foregoing description of the present invention is meant to be illustrative only. While a few examples of the present invention have been described in detail, the principles of the present invention may be adapted to many different variations without departing from embodiments of the invention.

[0071] According to a first embodiment there is provided a computerized method of generating an information display from an input of publication files containing text, graphics, and other data viewable as page images of a publication having stories (text passages) and graphics images appearing therein, comprising the steps of: extracting text data from the publication files corresponding to stories appearing in the page images of the publication, and maintaining them as text data files; processing page images from the publication files and maintaining them as page image files; mapping story areas for respective stories appearing in the page images and indexing each story area to a text data file corresponding to the text passage in the story area, and maintaining the mapped story areas as image map files; and generating a display on a computer system of page images using the page image files, and linking the stories in the story areas of the displayed page images to the corresponding text data using the text data files and image map files.

[0072] According to a second embodiment there is provided a computerized method of generating an information display from an input of publication files containing text, graphics, and other data viewable as page images of a publication having stories (text passages) and graphics images appearing therein, comprising the steps of: extracting text data from the publication files corresponding to stories appearing in the page images of the publication, and maintaining them as text data files; parsing the text data to find predetermined key-

words appearing therein, indexing each keyword to a page number and a story number for the story corresponding to the text passage in which the keyword is found, and maintaining the indexed keywords on a keyword list; processing page images from the publication files and maintaining them as page image files; generating a display on a computer system of the keyword list, and displaying the page image containing the story in which a selected keyword appears when the keyword is selected from the keyword list.

[0073] According to a third embodiment there is provided a computerized method of generating an information display from an input of publication files containing text, graphics, and other data viewable as page images of a publication having stories (text passages), and graphics images appearing therein, comprising the steps of: extracting text data from the publication files corresponding to stories appearing in the page images of the publication, and maintaining them as text data files; processing page images from the publication files and maintaining them as page image files; assigning to each story appearing on a page of the publication a page number on which the story appears, and a story number ranking corresponding to the relative importance of the story to other stories on the page; indexing the text data files to the page numbers and story number rankings for the corresponding stories appearing in the page images of the publication; generating a display on a computer system of a page image using the page image files, and a side-by-side display of a list of story titles for the stories appearing on the displayed page ranked in order of their assigned story number rankings.

Claims

1. A method of converting digital publication files containing text, graphics, and other data corresponding to pages of a publication having stories (text passages) and graphics images appearing therein to digital data for generating and controlling a display on a computer system, comprising the steps of:

extracting text data from the publication files corresponding to stories appearing in the pages of the publication, and maintaining them as digital text data

files; producing page images corresponding to pages of the publication and maintaining them as digital page image files;

mapping story areas for respective stories appearing in the pages of the publication and indexing each story area to a text data file corresponding to the text passage in the story area, and maintaining the mapped story areas as digital control data; and

producing from the digital text data files, the digital page image files and the digital control

data, digital data for generating a display on a computer system of page images in dependence on the page image files, and linking the stories in the mapped story areas of the displayed page images to the corresponding text data in dependence on the text data files and digital control data.

2. A method according to Claim 1, wherein said mapping step further includes the step of assigning to each mapped story area a page number of the page on which the story appears and a story number which corresponds to the relative importance of the story to other stories on the page.

3. A method according to Claim 1 or 2, wherein the story number is derived based upon any one of the following group of story importance indicators: location of the story area on the page; size of type font of a headline associated with the story; size of type font associated with the text passage corresponding to the story area; and size of text content of the text passage corresponding to the story area.

4. A method according to any preceding Claim, further comprising the steps of:

parsing the text data for the text passages to find predetermined keywords therein, indexing each keyword to a page number and a story number for the story corresponding to the text passage in which the keyword is found, maintaining the indexed keywords on a keyword list from which the corresponding story and text passage can be found, and using said keyword list in the production of said digital data.

5. A method according to Claim 4, wherein said parsing step is carried out automatically by performing a text string search of the text passages based upon text string entries contained in a library list of keywords.

6. A method according to any preceding Claim, further comprising the step of mapping graphics image areas for respective graphic images appearing in the page images and indexing each mapped graphics image to a page number of the page on which the graphics image appears and a graphics image area number, and maintaining the mapped graphics image areas as image map files.

7. A method according to any preceding claim, further comprising the step of generating a display on a computer system of page images in dependence on the page image files, and linking the stories in the mapped story areas of the displayed page images to the corresponding text data in depend-

ence on the text data files and digital control data.

8. A method of generating an information display according to Claim 7, wherein in the display generating step, a page image is displayed side-by-side with a text passage corresponding to a selected story appearing in the page image, whereby a viewer can read the text in the text passage while referring to the page image for visual cues about the text passage.
9. A method of generating an information display according to Claim 8, wherein the text passage is retrieved for simultaneous display with the page image upon clicking with a cursor on the corresponding story area defined in the image map for the story contained in the page image.
10. A method of generating an information display according to Claim 7 when dependent upon Claim 2, wherein the display generating step includes displaying a list of story titles ordered by page number and story numbers for stories appearing on a page simultaneously with display of the page image for the page.
11. A method of generating an information display according to Claim 7 when dependent upon Claim 4, wherein the display generating step includes displaying a list of keywords indexed in the publication, and displaying a text passage indexed to a keyword selected from the list of keywords simultaneously with the page image corresponding to the page on which the story corresponding to the text passage appears.
12. A method of generating an information display according to Claim 7 when dependent upon Claim 6, wherein the display generating step includes displaying further information linked to a graphics image appearing in a page simultaneously with the page image for the page in which the graphics image appears.
13. A method of generating an information display according to Claim 12, wherein the displayed further information is obtained from one of the following group of information sources: a stored graphics image; an output of an externally run process; and a stored text passage.
14. A method of converting digital publication files containing text, graphics, and other data corresponding to pages of a publication having stories (text passages) and graphics images appearing therein to digital data for generating and controlling a display on a computer screen, comprising the steps of:
 - extracting text data from the publication files corresponding to stories appearing in the page images of the publication, and maintaining them as digital text data files;
 - parsing the text data to find predetermined keywords appearing therein, indexing each keyword to a page number and a story number for the story corresponding to the text passage in which the keyword is found, and maintaining the indexed keywords as digital control data;
 - producing page images corresponding to the pages of the publication and maintaining them as digital page image files;
 - producing from the digital text data files, the digital page image files and the digital control data, digital data for generating a display on a computer system listing the keywords and linking each keyword to its corresponding story in dependence on said digital control data wherein on selection of a keyword from the list, the page image containing the story in which a selected keyword appears is displayed.
15. A method according to Claim 14, wherein the parsing step is carried out automatically by performing a text string search of the text passages based upon text string entries contained in a library list of keywords.
16. A method according to Claim 14, wherein the text data extracted from the publication files include type font delimiters identifying keywords appearing in the text passages, and the parsing step is carried out automatically by performing a search for keywords based upon locating them by their type font delimiters.
17. A method according to any one of claims 14, 15 or 16, further comprising the step of generating a display on a computer system listing the keywords and linking each keyword to its corresponding story in dependence on said digital control data wherein on selection of a keyword from the list, the page image containing the story in which a selected keyword appears is displayed.
18. A method of converting digital publication files containing text, graphics, and other data corresponding to pages of a publication having stories (text passages) and graphics images appearing therein to digital data for generating and controlling a display on a computer screen, comprising the steps of:
 - extracting text data from the publication files corresponding to stories appearing in the page images of the publication, and maintaining them as text data files;
 - producing page images corresponding to the

pages of the publication and maintaining them as page image files;

assigning to each story appearing on a page of the publication a page number on which the story appears, and a story number ranking corresponding to the relative importance of the story to other stories on the page;

indexing the text data files to the page numbers and story number rankings for the corresponding stories appearing in the page images of the publication;

producing from the digital text data files and the page image files, digital data for generating a display on a computer system of a page image in dependence on the page image files, and, in dependence on said text data files and indexing, a side-by-side display of a list of story titles for the stories appearing on the displayed page ranked in order of their assigned story number rankings.

19. A method according to Claim 18, wherein the story number rankings are derived based upon comparing any one or more of the following group of story importance indicators: location of the story area on the page; size of type font of a headline associated with the story; size of type font associated with the text passage corresponding to the story area; and size of text content of the text passage corresponding to the story area.
20. A method according to claim 18 or 19, further comprising the step of generating a display on a computer system of a page image in dependence on the page image files, and, in dependence on said text data files and indexing, a side-by-side display of a list of story titles for the stories appearing on the displayed page ranked in order of their assigned story number rankings.
21. A method according to Claim 20, wherein the text data for a text passage corresponding to a selected story is retrieved using the text data files upon clicking with a cursor on a selected story title on the list of story titles.
22. A method according to Claim 20 or 21, wherein the display generating step includes displaying the text passage corresponding to a selected story title simultaneously with the page image for the page in which the selected story appears, whereby a viewer can read the text in the text passage while referring to the page image for visual cues about the text passage.
23. A method according to any one of Claims 18 to 22, further comprising the steps of:

parsing the text data to find predetermined keywords appearing therein, indexing each keyword to a page number and a story number for the story corresponding to the text passage in which the keyword is found, and maintaining the indexed keywords on a keyword list;

generating a display on a computer system of the keyword list, and simultaneously displaying the page image containing the story in which a selected keyword appears side-by-side with the text passage in which selected keyword appears, when the keyword is selected from the keyword list.

24. A digital data structure, for remote reproduction of a publication on a visual display unit of a computer system, comprising:

extracted text data for reproducing the textual content of stories appearing in the pages of the publication;

text indexed data indexing each story on each page of the publication to the extracted text data for reproducing the textual content of that story;

page image data for reproducing a visual image of pages of the publication;

image index files indexing each page of the publication to the page image data for reproducing that page;

control data including mapping data indexing mapped areas of the visual image of the publication, as defined by the page image data, to text portions defined by the extracted text data; and

an interfacing instruction sequence for execution by a computer system to generate a display of visual images of pages of the publication in dependence on the page image data and image index data, a display of the text of stories in the publication in dependence on the extracted text data and text index data, and providing for the linking of the mapped story areas on the visual images of pages of the publication to text portions in dependence on the control data and extracted text data, wherein selection of a first mapped story area effects the display of the text corresponding to that first mapped story area.

25. A digital data structure, for remote reproduction of a publication on a visual display unit of a computer system, comprising:

extracted text data for reproducing the textual content of stories appearing in the pages of the publication;

text index data indexing each story on each

page of the publication to the extracted text data for reproducing the textual content of that story;

page image files for reproducing the visual images of pages of the publication;

image index data indexing each page of the publication to the page image data for reproducing that page;

keyword data for reproducing keywords appearing in the pages of the publication;

keyword index data indexing each keyword to the story in which the keyword is found; and

an interfacing instruction sequence for execution by a computer system to generate a display of a list of keywords and provide for linking each keyword to its corresponding story wherein on selection of the keyword from the list the page image containing the story in which the selected keyword appears is displayed.

26. A digital data structure, for remote reproduction of a publication on a visual display unit of a computer system, comprising:

extracted text data for reproducing the textual content of stories appearing in the pages of the publication;

text index data indexing each story on each page of the publication to the extracted text data for reproducing the textual content of that story, and ranking the stories appearing on a page of the publication relative to one another;

page image files for reproducing a visual images of pages of the publication;

image index data indexing each page of the publication to the page image data for reproducing that page; and

an interfacing instruction sequence for execution by a computer system to generate a display of visual images of pages of the publication in dependence on the page image data and image index data, and, in dependence on said extracted text data and text index data, a side-by-side display of a list of story titles for the stories appearing on the displayed page ranked in order of their assigned story number rankings.

27. An information display system comprising:

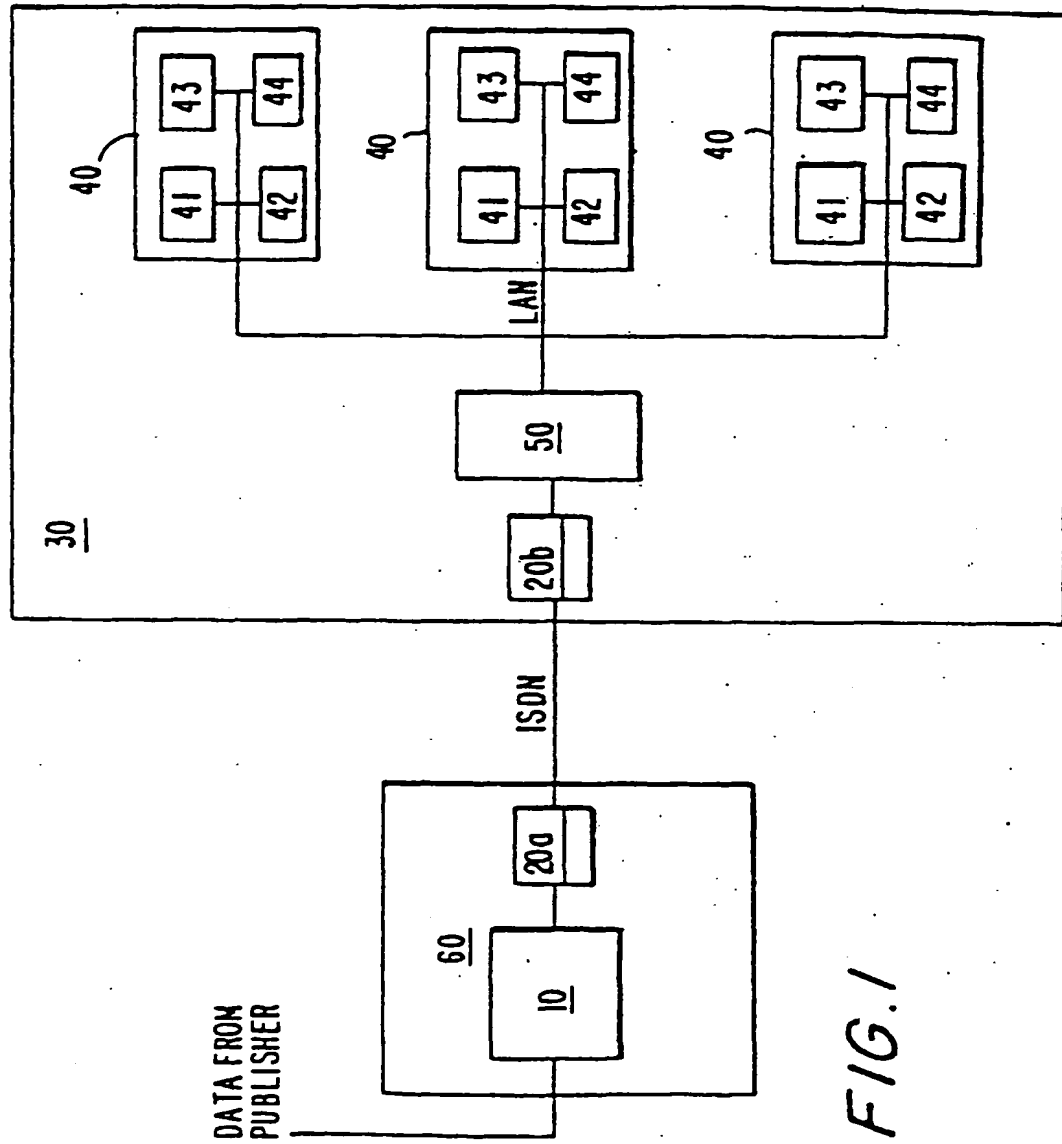
data access means to access page image data comprising graphical images of pages of a publication and text data, said page image data comprising graphical images of both text and non-text matter, said text matter including a plurality of predefined passages of text which are arranged on the page in an ordering indicating

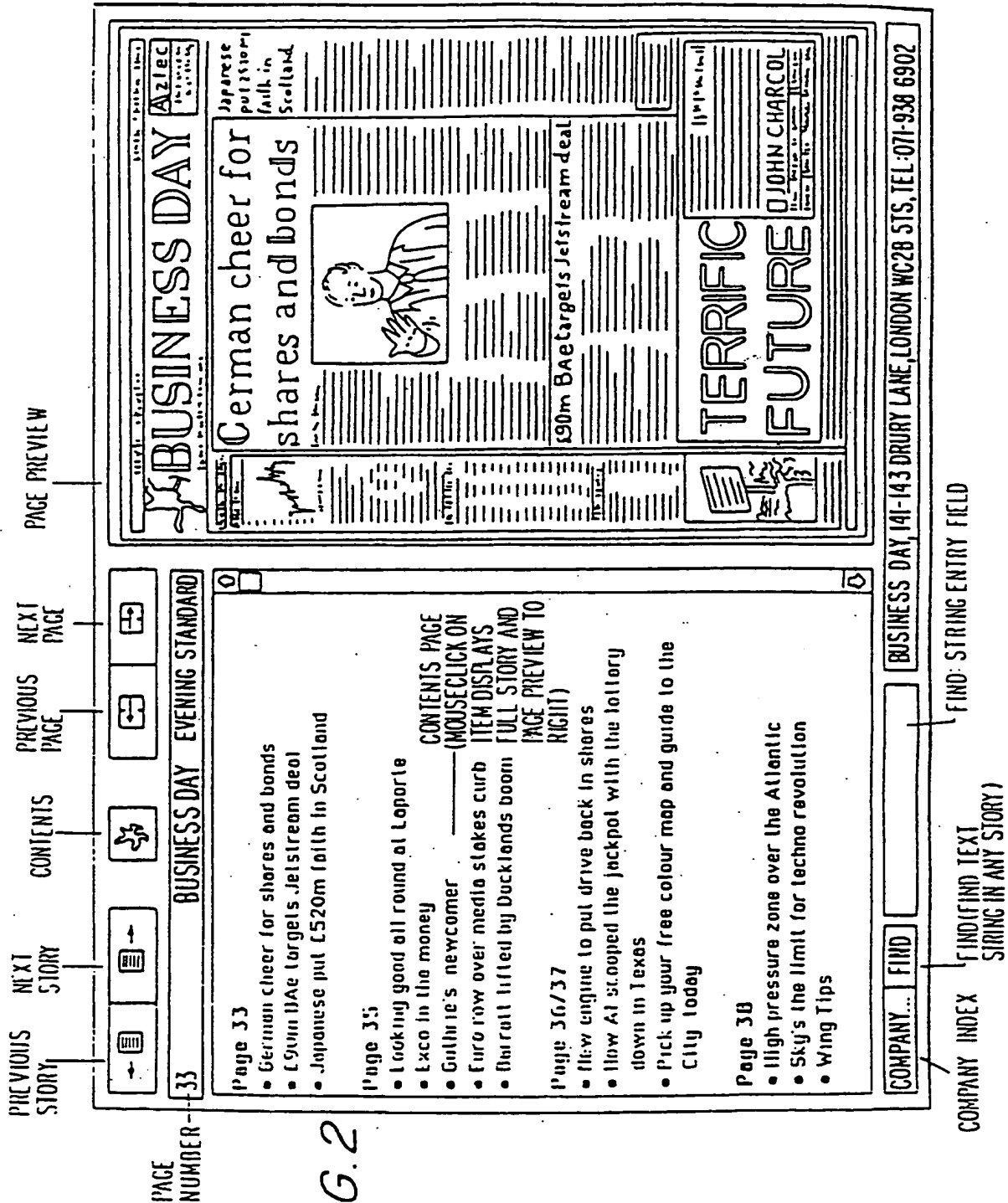
their relative importance in relation to other text passages on the page, and said text data comprising a plurality of predefined passages of text corresponding to text passages appearing in the graphical images of the pages of the publication;

display means to display a graphical image of a page of the publication simultaneously with display of text data corresponding to a text passage appearing in the graphical image of the displayed page; and

selection means operable by a user to select a text passage appearing in the graphical image of the page displayed by said display means;

said display means being responsive to said selection means and said data access means to simultaneously display the graphical image of the page of the publication accessed from said page image data alongside the selected passage of text accessed from said text data such that the user can view the text passage in detail as text data while being apprised of its relative importance in relation to the other text passages through viewing the simultaneous display of the graphical image of the pages and wherein said display means is adapted to visually identify said selected portion of the displayed page of said images including said predefined passage of text and any associated non-text matter.





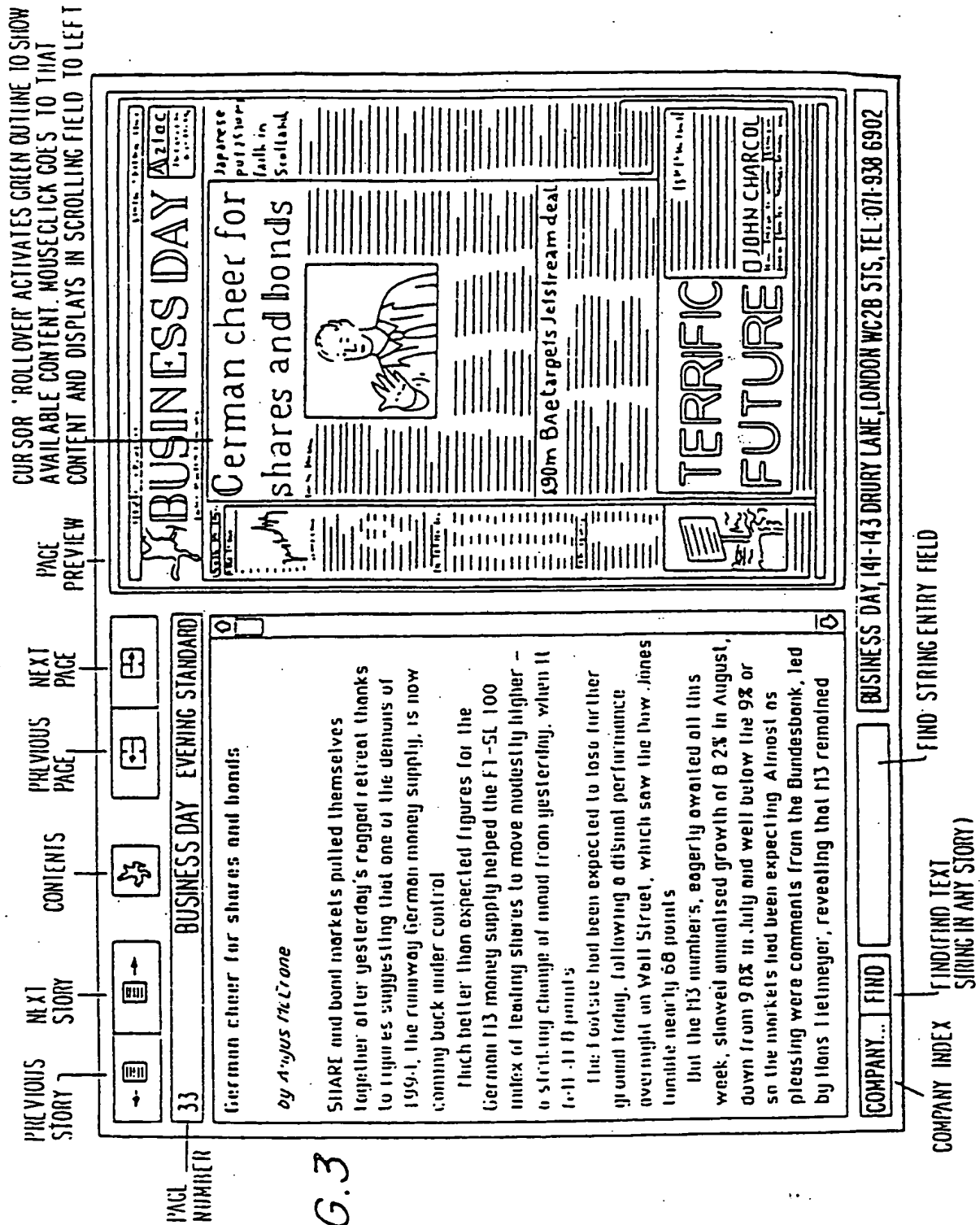
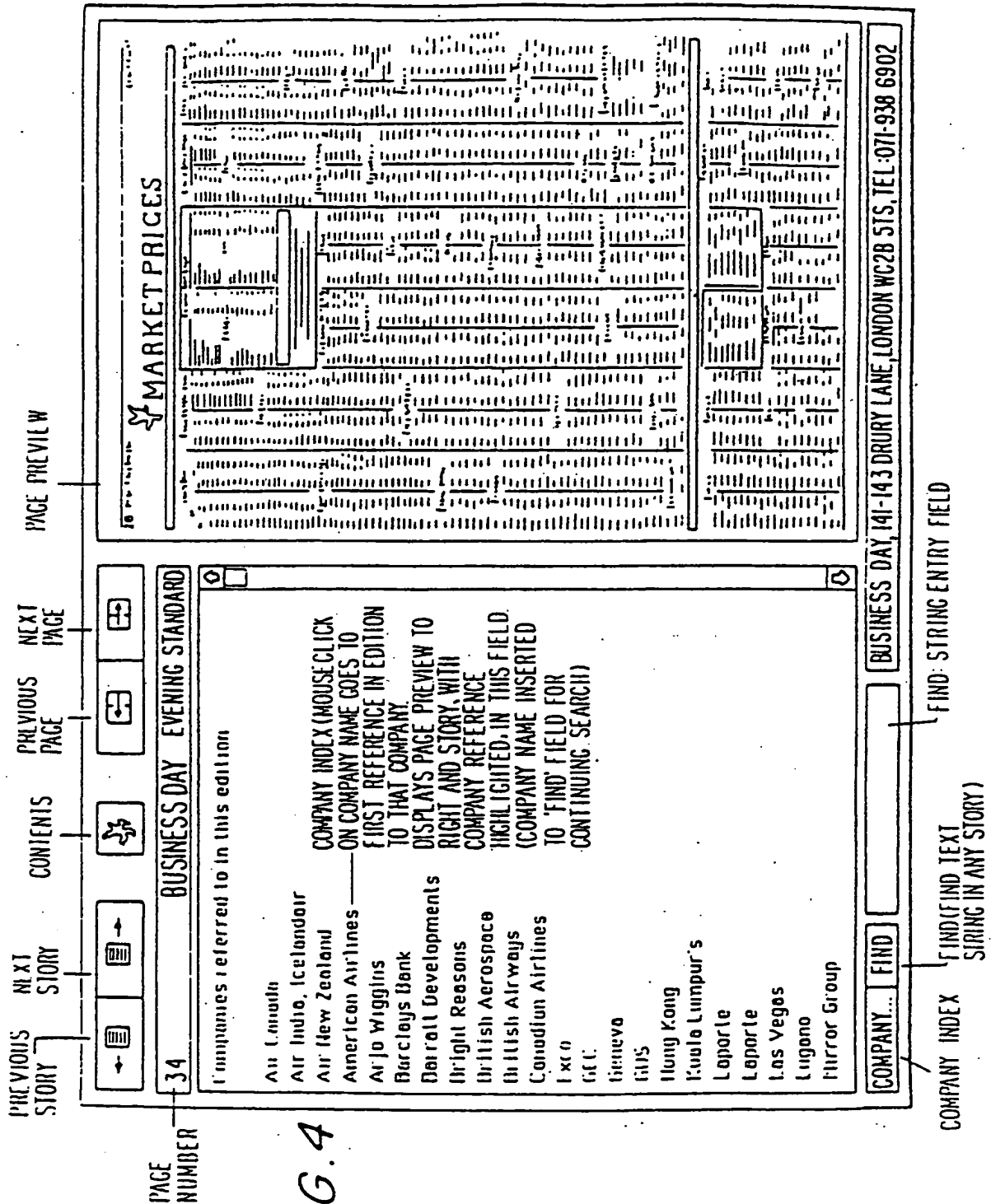


FIG. 3



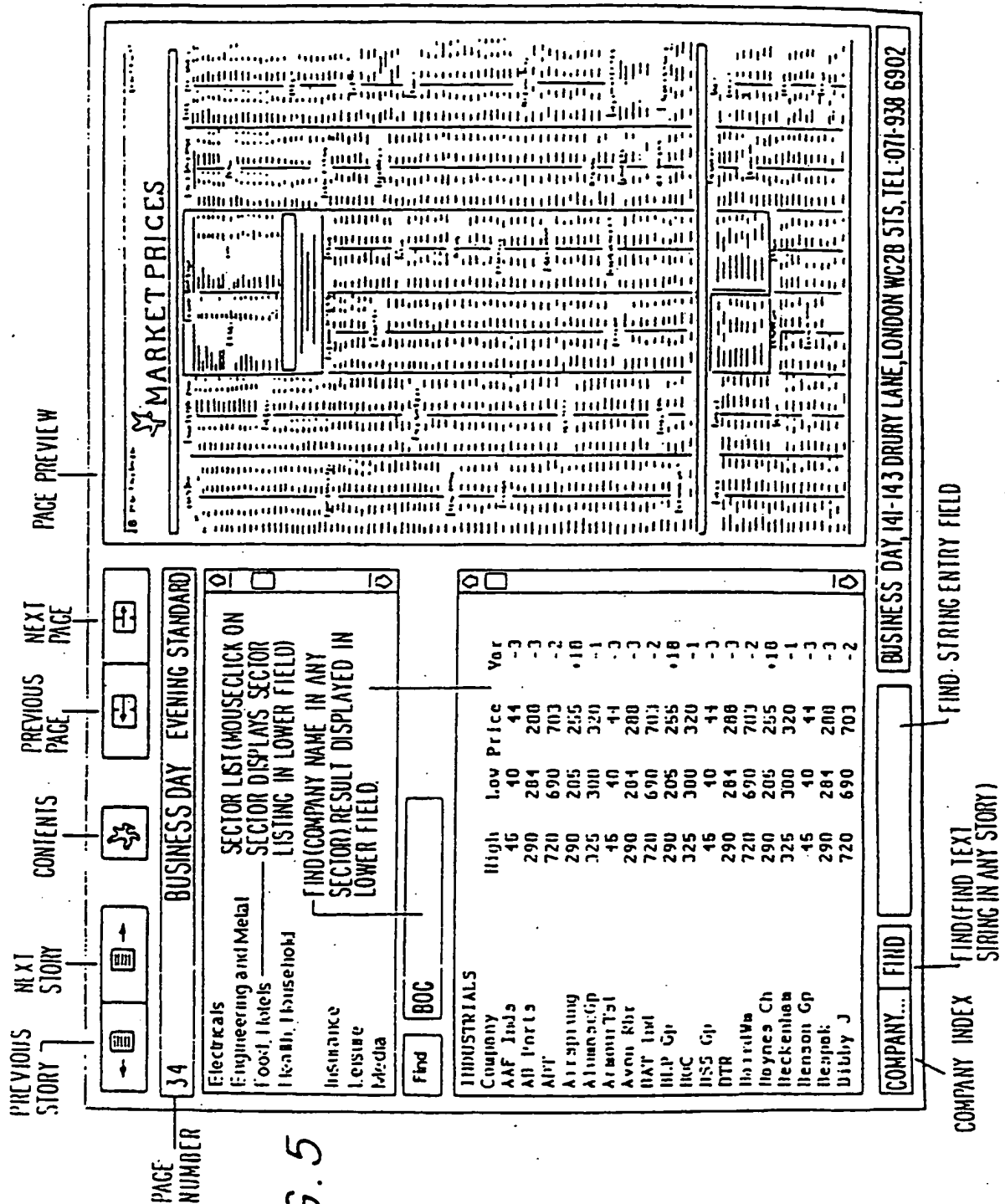


FIG.6

PAGE NUMBER 39 BUSINESS DAY EVENING STANDARD

PREVIOUS STORY NEXT STORY PREVIOUS PAGE NEXT PAGE CONTENTS

PAGE PREVIEW

Footsie reels from Iraqi shockwaves

Bondelsok FALL
to temple
buyers skilful

APPEALON M.L. (M.L.)

Stanley have also turned negative on the stock, cutting their earnings forecast to 46p from 51p a share for 1996 and to 41p from 55p for 1997 on the back of a forecast fall in sales of anti-ulcer drug Zentac.

They have cut their sales estimate to £2.056 billion from £2.270 billion for 1996 and are going for £1.7 billion for 1997, against previous hopes of £2.2 billion.

Other pharmaceuticals stocks were also weak. Wellcome lost 11p to 541p. SmithKline Beecham was a penny lower at 364p and Zeneca slipped 9p to 725p.

Housebuilder Persimmon was steady at 289p as CL-Teag placed the rump of the rights issued.

Brickmaker Ibstock Johnson

Johnson reports result next week, which should show that the British side has turned up sharply.

Higher volumes and greater utilisation rates will push profits ahead, although the overall result will be dragged down by pulp, say MetWest Markets.

The pulp price touched its lowest level for 30 years in real terms in 1993 and a 10% rise would add £2 million to the bottom line.

price 95 90 85 80 75

Jan Feb Mar Apr

COMPANY... FIND

FIND(FIND TEXT STRING IN ANY STORY)

FIND STRING ENTRY FIELD

BUSINESS DAY, 141-143 DRURY LANE, LONDON WC2B 8TS, TEL: 071-938 6902

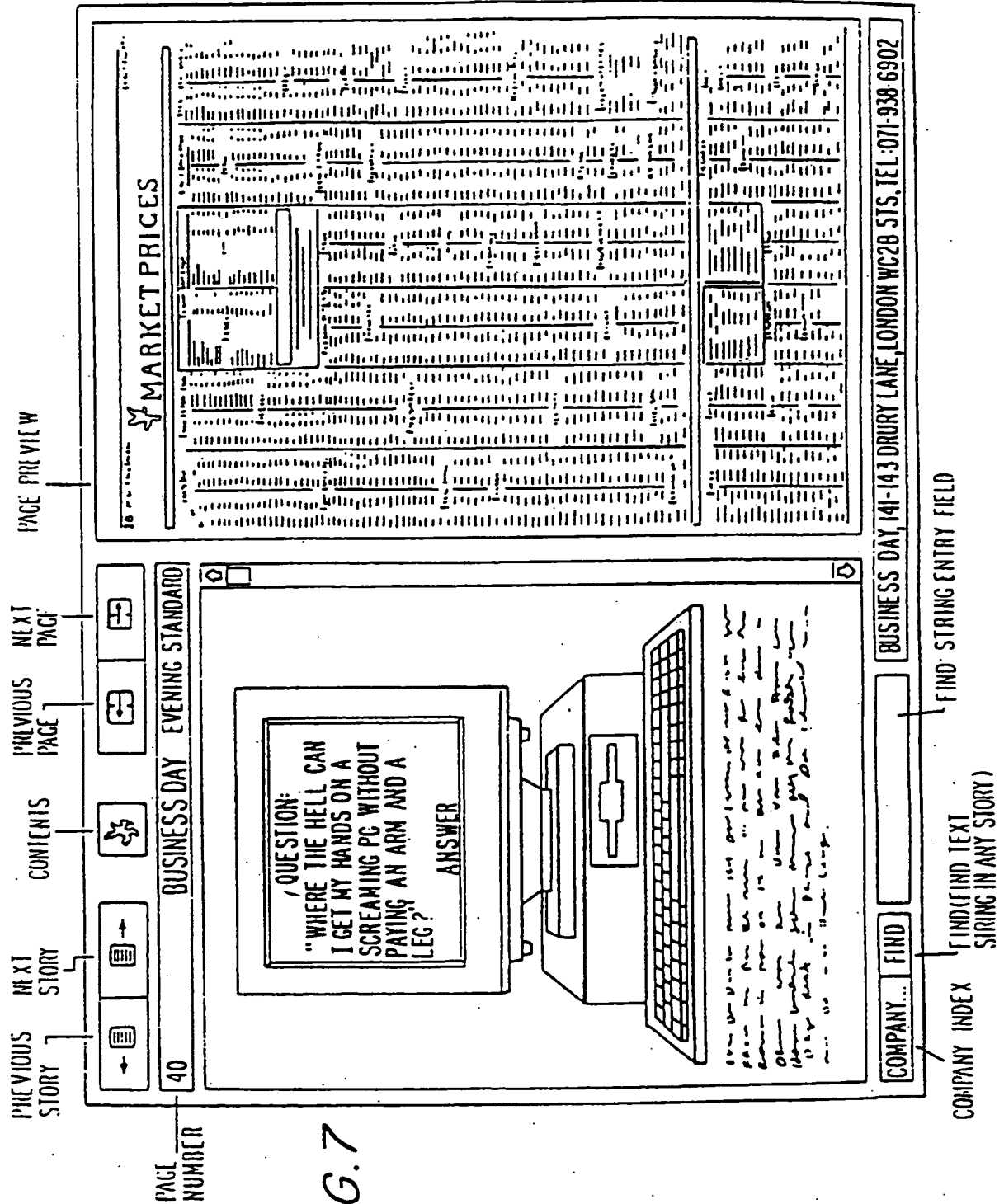
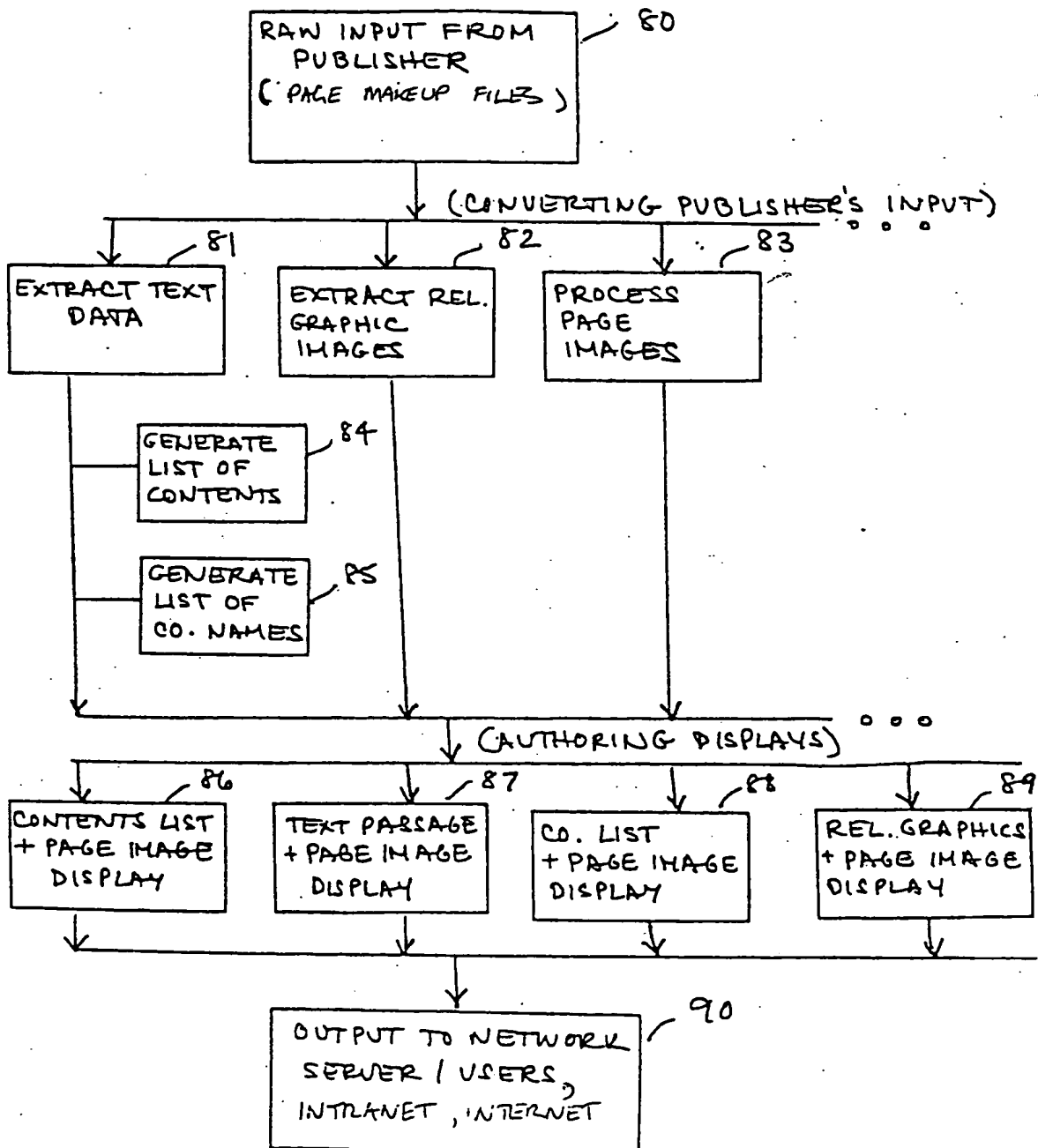


Fig. 8



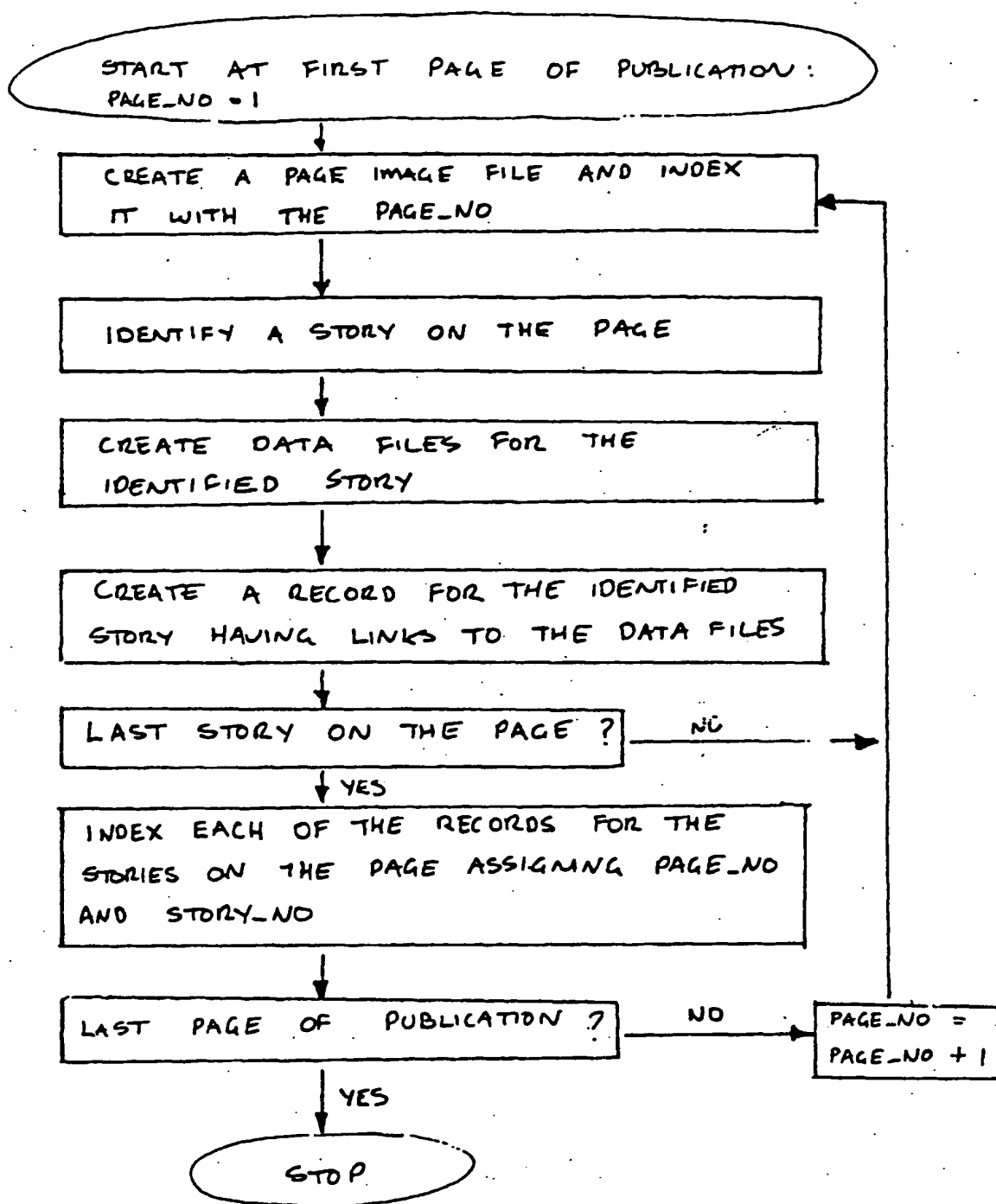


Fig 9a.

Fig. 9b.

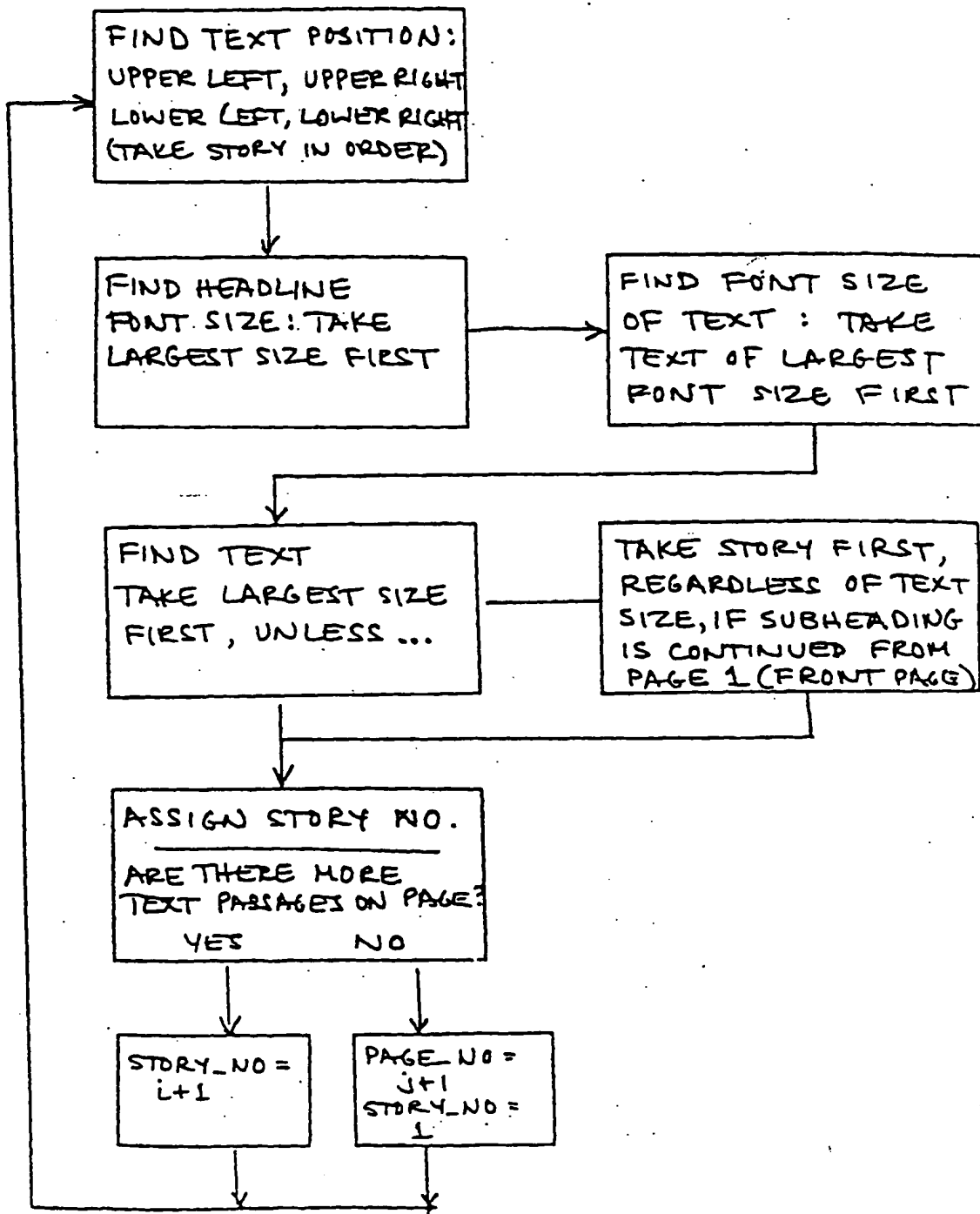


Fig. 10

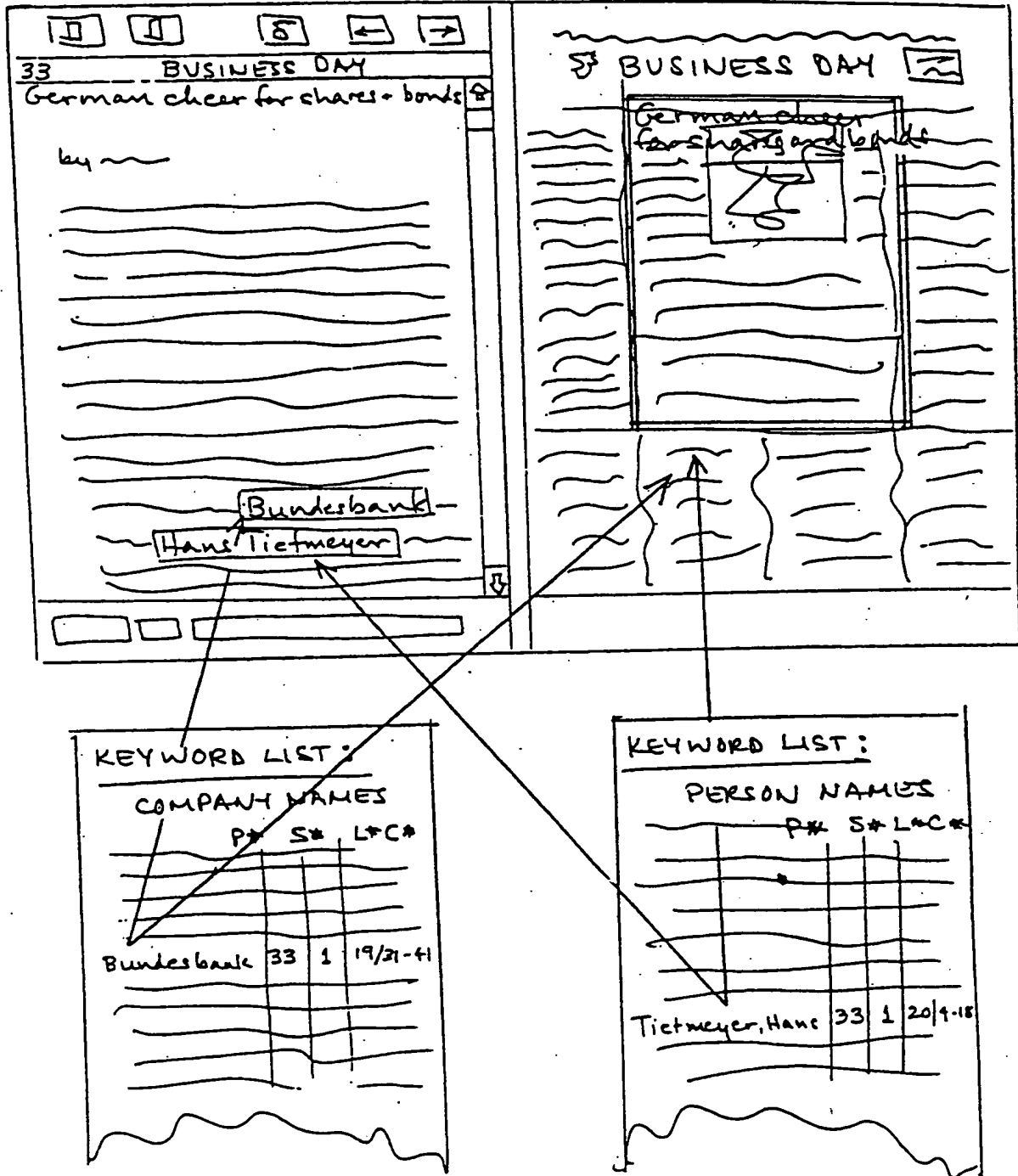
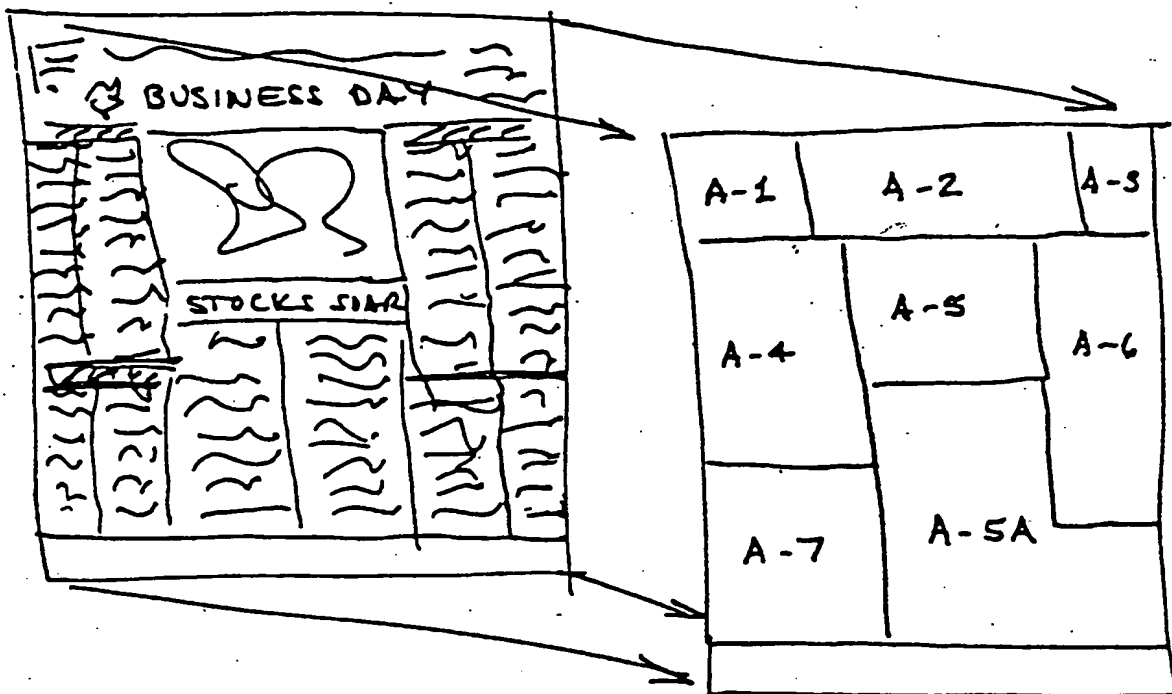
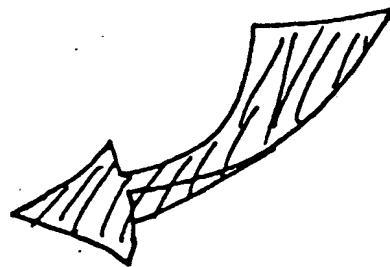


Fig. 11



AREA	STORY / PIECE	TEXT LOC.
1	(Weather)	0100...
2	(Headline)	0150...
3	(Index to Issue)	0180...
4	Story # 1	1010...
5	Photo, Story # 2	2050...
5A	Story # 2	2010...
6	Story # 3	3010...
7	Story # 4	4010...
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮





European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 97 31 0368

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	US 5 524 201 A (SHWARTS SCOTT L ET AL) 4 June 1996 * abstract * * column 10, line 43 - column 12, line 49 *	1-27	G06F17/60
X	US 5 502 637 A (BEAULIEU ROLAND C ET AL) 26 March 1996 * column 4, line 55 - column 5, line 15 *	1	
A	WO 93 15466 A (RAYCHEM) 5 August 1993 * page 31, line 28 - line 35 * * page 38, line 12 - line 17 *	1-27	
A	US 5 675 788 A (HUSICK LAWRENCE A ET AL) 7 October 1997 * column 26, line 26 - column 30, line 47 *	1-27	
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)
			G06F H04L
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 25 May 1998	Examiner Pottiez, M
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document</p>			

EPO FORM 1503 03 82 (P4C01)